

## ANALISIS SELEKSI ATRIBUT DALAM MEMPREDIKSI KEGAGALAN SKRIPSI MAHASISWA MENGGUNAKAN ALGORITME NAÏVE BAYES PADA JURUSAN SISTEM INFORMASI UNIVERSITAS BRAWIJAYA

Satrio Hadi Wijoyo<sup>\*1</sup>, Satrio Agung Wicaksono<sup>2</sup>, Admaja Dwi Herlambang<sup>3</sup>

<sup>1,2,3</sup>Universitas Brawijaya, Malang

Email: <sup>1</sup>satriohadi@ub.ac.id, <sup>2</sup>satrio@ub.ac.id, <sup>3</sup>herlambang@ub.ac.id

\*Penulis Korespondensi

(Naskah masuk: 14 Juli 2022, diterima untuk diterbitkan: 11 April 2023)

### Abstrak

Data mining dapat diterapkan pada bidang lembaga atau institusi Pendidikan dan sering disebut juga dengan *Educational Data Mining* (EDM) yaitu sebuah pengembangan metode dalam mengeksplorasi jenis tipe data pendidikan yang bersifat unik yang bertujuan untuk mempelajari dalam memahami kinerja siswa dan pengaturan lingkungan di tempat siswa belajar. Data mining juga dapat dimanfaatkan untuk analisa seleksi atribut untuk prediksi kegagalan belajar mahasiswa pada perkuliahan. Salah satunya kegagalan dalam menyelesaikan skripsi dengan menggunakan algoritme naïve bayes. Tahapan penelitian ini dimulai dari pengumpulan data, preproses data, implementasi algoritme naïve bayes, pengujian, dan analisa hasil. Data yang digunakan sebanyak 500 data untuk mahasiswa angkatan 2012-2015 dengan 24 atribut. Akurasi tertinggi yaitu algoritme naïve bayes dengan menggunakan data latih yang menunjukkan hasil 412 prediksi benar dengan akurasi sebesar 82.4% dan 88 prediksi salah dengan presentasi sebesar 17.6%. Sedangkan hasil seleksi atribut yang telah dilakukan terdapat Lama P2 merupakan atribut dengan ranking teratas yang mempengaruhi hasil klasifikasi.

**Kata Kunci :** *Algoritme Naïve Bayes, Kegagalan Skripsi Seleksi Atribut*

## ANALYSIS OF ATTRIBUTES SELECTION IN PREDICTING STUDENT'S THESIS FAILURE USING NAÏVE BAYES ALGORITHM IN THE DEPARTMENT OF INFORMATION SYSTEMS UNIVERSITY OF BRAWIJAYA

### Abstract

Data mining can be applied to the field of educational institutions or institutions and is often referred to as *Educational Data Mining* (EDM), which is a method development in exploring unique types of educational data types that aim to study in understanding student performance and environmental settings in which students learn. Data mining can also be used to analyze attribute selection for predicting student learning failure in lectures. One of them is the failure in completing the thesis using the naïve Bayes algorithm. The stages of this research started from data collection, data preprocessing, implementation of the Naïve Bayes algorithm, testing, and analysis of the results. The data used is as much as 500 data for students class 2012-2015 with 24 attributes. The highest accuracy is the Naive Bayes algorithm using data training which shows the results of 412 correct predictions with an accuracy of 82.4% and 88 false predictions with a presentation of 17.6%. Meanwhile, the result of attribute selection that has been carried out is that the Old P2 is the attribute with the top ranking that affects the classification results.

**Keywords:** *Naïve Bayes Algorithm, Attribute Selection Thesis Failure*

### 1. PENDAHULUAN

Data Mining sudah banyak diterapkan pada berbagai bidang, misalnya bidang analisis pemasaran, kedokteran, pendidikan, dan lain-lain. Data mining dapat diterapkan pada bidang lembaga atau institusi Pendidikan dan sering disebut juga dengan *Educational Data Mining* (EDM) yaitu sebuah pengembangan metode dalam mengeksplorasi jenis tipe data pendidikan yang bersifat unik yang

bertujuan untuk mempelajari dalam memahami kinerja siswa dan pengaturan lingkungan di tempat siswa belajar (Romero & Ventura, 2010). EDM dapat diterapkan dengan banyak metode seperti menggunakan teknik Pohon Keputusan (*Decision Tree*), Jaringan Syaraf Tiruan, *Naive Bayes Classifier*, dan lain-lain.

Salah satunya adalah algoritme Naive Bayes yang sudah sering digunakan dalam penelitian di

bidang pendidikan. *Teorema Bayes* diterapkan pada algoritme *naïve bayes* dengan mengasumsikan semua fitur saling tidak berhubungan. Algoritme *naïve bayes* menganggap bahwa ada atau tidaknya fitur tertentu dari kelas tidak berhubungan dengan ada tidaknya fitur yang lain (Yuan, 2010). Algoritme *Naive Bayes* memiliki proses komputasi yang sederhana dan akurasi yang tinggi.

Penelitian terkait penerapan data mining di bidang pendidikan telah dilakukan untuk mengklasifikasikan judul skripsi dengan menggunakan algoritme *naïve bayes classifier* dan algoritme pohon keputusan (*decision tree*). Hasil penelitian tersebut ditemukan hasil perbandingan akurasi yang signifikan dengan akurasi 80,33% untuk algoritme *naïve bayes* dan 60,33% untuk algoritme pohon keputusan (Nuraeni, R., dkk. 2021). Klasifikasi media sosial terkait komentar positif dan negatif menggunakan perbandingan algoritme *Naive Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine*. Algoritme tersebut mengklasifikasikan data kalimat sebanyak 2000 data. Hasil penelitian tersebut memiliki nilai akurasi sebesar 79,8% untuk algoritme *Naive Bayes*, nilai akurasi sebesar 50,23% untuk algoritme *K-Nearest Neighbor*, dan nilai akurasi sebesar 75,29% untuk *Support Vector Machine* (Asnawi, M. H., dkk. 2021). Klasifikasi mahasiswa putus sekolah dengan menggunakan empat model klasifikasi yaitu *Naive Bayes Classifier*, *Random Forest*, *J48 graft* dan *Bayessian Logistic*. Tujuan penelitian tersebut untuk membantu para mahasiswa dan administrasi akademik dalam melakukan identifikasi dan perbaikan agar mahasiswa dapat memperbaiki kinerjanya dan membantu mengurangi mahasiswa putus sekolah (*Drop Out*). Hasil evaluasi algoritme *Naive Bayes Classifier* memiliki tingkat akurasi tertinggi dibandingkan dengan tiga algoritme lainnya sebesar 91,9355% (Shakeel, K., & Anwer Butt, N., 2015).

Klasifikasi materi teks dari website untuk mata pelajaran produktif. Tujuan penelitian tersebut untuk mendeskripsikan hasil evaluasi klasifikasi pada dokumen materi teks dengan menggunakan algoritme *Naive Bayes*. Nilai akurasi tertinggi sebesar 81,48% (Herlambang, A., D., dan Satrio H., W., 2019). Prediksi dilakukan pada siswa SMK berdasarkan faktor internal dan faktor eksternal. Prediksi tersebut menggunakan algoritme *naïve bayes*. Hasil evaluasi *naïve bayes* sebesar 83% (Herlambang, A., D., dkk, 2019). Klasifikasi kemampuan siswa berdasarkan kemampuan intelektual dan faktor eksternal pada siswa SMA Tuna Luluh menggunakan *naïve bayes*. Data yang digunakan sebanyak 190 data siswa. Hasil pengujian menunjukkan nilai akurasi sebesar 92,86%, presisi sebesar 94,73%, recall sebesar 94,73%, dan *F-Measure* sebesar 96% (Wicaksono, A. H., dkk, 2022). Hasil dari beberapa penelitian terdahulu menunjukkan bahwa algoritme *naïve bayes* terbukti unggul dalam pengujian. Sehingga penelitian ini menggunakan algoritme *naïve bayes* untuk

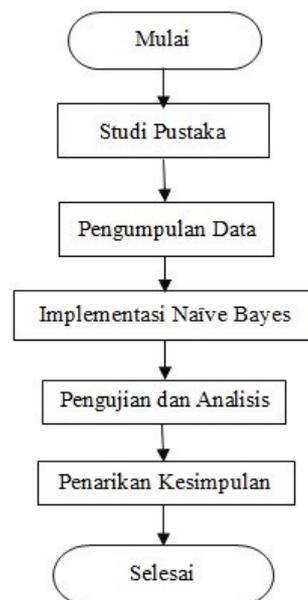
analisis seleksi atribut untuk memprediksi kegagalan skripsi.

Salah satunya kegagalan mahasiswa dalam menyelesaikan skripsinya adalah waktu pengerjaan, menunda-nunda pengerjaan, lama studi, nilai ipk atau ips kurang baik sehingga mahasiswa harus mengulang, dan faktor yang lainnya. Faktor-faktor tersebut yang akan diseleksi. Pada penelitian ini yang digunakan seperti nama, nilai mahasiswa setiap semester, lama pengerjaan skripsi pada setiap tahapan, dan kendala yang lainnya. Data sekunder diperoleh dari *database* fakultas ilmu komputer. Pencarian faktor yang berpengaruh dalam memprediksi kemungkinan mahasiswa mengalami kegagalan penyelesaian skripsi. Jika faktor yang berpengaruh sudah diketahui maka penelitian ini dapat memberikan rekomendasi kepada fakultas ilmu komputer.

Berdasarkan permasalahan dan dukungan penelitian sebelumnya terkait penggunaan algoritme *Naive Bayes*. Tujuan penelitian ini untuk menganalisis seleksi atribut dalam memprediksi kegagalan skripsi mahasiswa menggunakan algoritme *Naive Bayes* pada Jurusan Sistem Informasi Universitas Brawijaya.

## 2. METODE PENELITIAN

Tahapan ini dibahas metode penelitian yang digunakan dan langkah-langkah yang dilakukan dalam penelitian mengenai analisis seleksi atribut dalam memprediksi kegagalan skripsi mahasiswa menggunakan algoritme *Naive Bayes* pada Jurusan Sistem Informasi Universitas Brawijaya. Tahapan penelitian ditunjukkan pada Gambar 1.



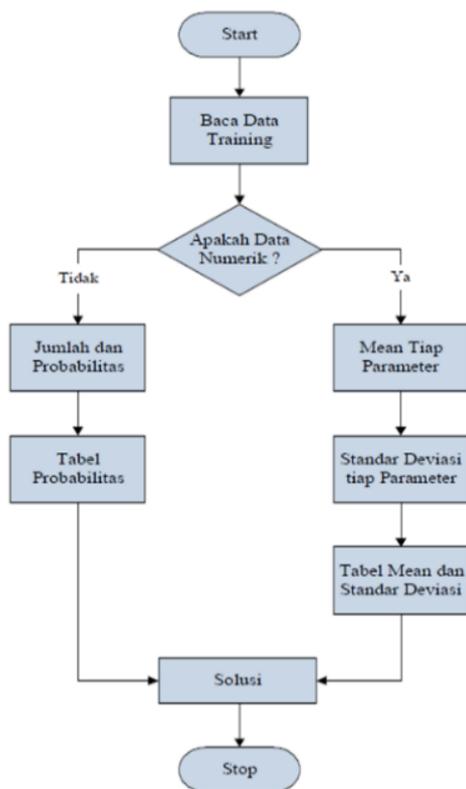
Gambar 1. Tahapan Penelitian

Studi literatur merupakan langkah awal dalam proses pengumpulan informasi sebagai acuan, studi literatur merupakan metode pengumpulan informasi melalui dokumen-dokumen baik berupa buku, jurnal

maupun artikel dari internet. Dalam penelitian ini, literatur yang dipelajari meliputi: analisis seleksi atribut-tribut, prediksi kegagalan skripsi mahasiswa. algoritme *Naïve Bayes*.

Pengumpulan data menggunakan data sekunder. Data sekunder diperoleh bagian akademik Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Brawijaya. Data tersebut memiliki beberapa atribut yang digunakan untuk menentukan mahasiswa mengalami kegagalan skripsi atau tidak. Data yang digunakan adalah tahun akademik 2019-2020. Jurusan sistem informasi memiliki tiga program studi, yaitu program studi sistem informasi, pendidikan teknologi informasi, dan teknologi. Data dari ketiga program studi tersebut akan digunakan pada penelitian ini.

Proses implementasi algoritme *Naive Bayes* merupakan proses implementasi menggunakan bahasa pemrograman Weka. *Flowchart* algoritme *Naive Bayes* dapat dilihat pada Gambar 2. Sebelum proses implementasi algoritme *Naive Bayes*, data yang diperoleh dilakukan tahapan *preprocessing*. Pada tahap *preprocessing* data, terdapat tiga proses yang dilakukan, yaitu proses seleksi data, pembersihan data dan transformasi data.



Gambar 2. Flowchart Algoritme Naive Bayes

Proses pengujian dan analisis dilakukan terhadap proses perhitungan dari algoritme yang telah diimplementasikan untuk melakukan analisis seleksi atribut dalam memprediksi kegagalan skripsi mahasiswa menggunakan algoritme *naïve bayes* pada jurusan sistem informasi universitas brawijaya dengan baik. Pengujian dilakukan dengan beberapa

skenario untuk mengetahui seberapa baik hasil optimasi yang dapat dilakukan oleh sistem. Analisis dilakukan untuk menentukan seberapa baik sistem bekerja dengan cara menentukan nilai akurasi paling baik sebagai representasi solusi yang dihasilkan oleh implementasi algoritme *naive bayes* dengan pemrograman Weka.

Penarikan kesimpulan diambil sebagai hasil akhir dari penelitian yang dilakukan. Kesimpulan dapat diambil setelah semua proses penelitian dilakukan mulai dari proses pengumpulan data hingga hasil analisis yang memaparkan hasil akhir yang didapatkan dari proses penelitian yang dilakukan. Selain itu dipaparkan pula saran sebagai perbaikan atas penelitian ini sebagai pertimbangan perbaikan untuk penelitian dan pengembangan selanjutnya.

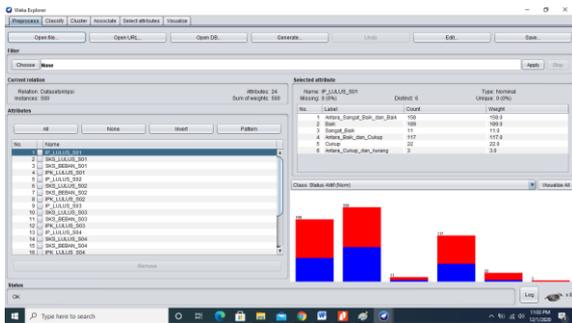
### 3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini diperoleh dari *database* Fakultas Ilmu Komputer melalui petugas akademik Fakultas Ilmu Komputer. Data yang didapatkan dari akademik berjumlah 1565 data yang terdiri dari data mahasiswa jurusan sistem informasi mulai angkatan 2009-2019. Data yang didapatkan dari akademik memiliki 54 atribut yaitu ID Mahasiswa, Jenis Kelamin, Jalur Masuk, Kota Asal, Asal SMA, Pekerjaan Ayah, Pendapatan Ayah, Pendidikan Ayah, Pekerjaan Ibu, Pendapatan Ibu, Pendidikan Ibu, Tanggal *upload* Praproposal, Tanggal *upload* Proposal, Tanggal pelaksanaan P0, Tanggal pelaksanaan P1, Tanggal pelaksanaan P2, Tanggal pelaksanaan Semhas, Tanggal pelaksanaan Ujian Skripsi, Indeks Prestasi Beban Semester 1-5, SKS Beban Semester 1-5, Indeks Prestasi Kumulatif Beban Semester 1-5, SKS Kumulatif Beban Semester 1-5, Indeks Prestasi Lulus Semester 1-5, SKS Lulus Semester 1-5, Indeks Prestasi Kumulatif Lulus Semester 1-5, SKS Kumulatif Lulus Semester 1-5, Indeks Prestasi Kumulatif Lulus, Predikat Kelulusan, dan Yudisium.

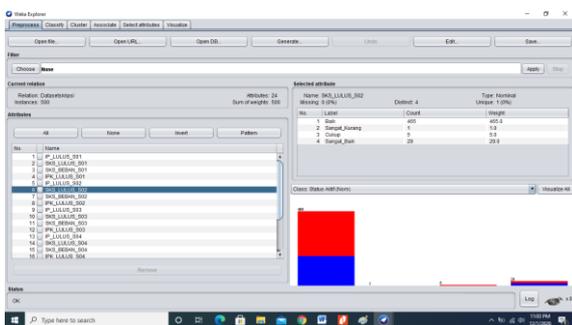
Hasil dari tahapan *preprocessing* data yang digunakan adalah data mahasiswa sistem informasi angkatan 2012-2015 yang sudah dinyatakan lulus atau selesai dan gagal skripsi. Data mahasiswa yang mengerjakan skripsi dibedakan menjadi 2, yaitu selesai skripsi dan gagal skripsi. Mahasiswa dapat dikatakan selesai skripsi apabila lama pengerjaan skripsi dari mahasiswa tersebut selama 1 semester. Data mahasiswa dibawah angkatan 2012 dihapus karena mahasiswa angkatan 2011 memiliki kurikulum yang berbeda. Mahasiswa angkatan 2016-2019 juga dihapus karena belum mengerjakan skripsi. Selain data-data yang tidak relevan, data mahasiswa dengan jalur masuk SAP juga dihilangkan karena data akademik yang dimiliki mahasiswa SAP berbeda dengan mahasiswa dari jalur masuk yang lain. Berdasarkan data yang didapat dari database Fakultas Ilmu Komputer, terdapat 500 data untuk mahasiswa angkatan 2012-2015 yang dilakukan tahapan

implementasi Algoritmes *Naive Bayes*. Dan atribut data mahasiswa yang digunakan 24 atribut yaitu Indeks Prestasi Beban Semester 1-5, SKS Beban Semester 1-5, Indeks Prestasi Kumulatif Beban Semester 1-5, SKS Kumulatif Beban Semester 1-5, Indeks Prestasi Lulus Semester 1-5, SKS Lulus Semester 1-5, Indeks Prestasi Kumulatif Lulus Semester 1-5, SKS Kumulatif Lulus Semester 1-5, Tanggal upload Proposal, Tanggal pelaksanaan P0, Tanggal pelaksanaan P1, dan Tanggal pelaksanaan P2. Tanggal tahapan pengerjaan skripsi dihitung menjadi lama dalam mengerjakannya.

Hasil implementasi algoritme *Naive Bayes* menggunakan Weka mulai dari hasil visualisasi atribut data sampai hasil seleksi atribut. Bentuk visualisasi atribut menggunakan tools weka dari beberapa atribut dapat dilihat pada Gambar 3, 4, 5, dan 6. Gambar 3 menunjukkan hasil visualisasi dari atribut IP Lulusan Semester 1. Nilai IP yang berupa data numerik ditransformasikan menjadi 6 kategori. Diketahui dari 500 data mahasiswa, ada 11 kategori Sangat Baik, 158 kategori Antara Sangat Baik dan Baik, 189 kategori Baik, 117 kategori antara baik dan cukup, 22 kategori cukup, 3 kategori antara cukup dan kurang, dan 0 kategori kurang.

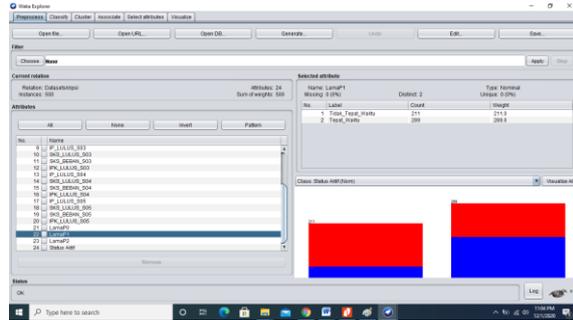


Gambar 3. Visualisasi dari IP Lulus Semester 1

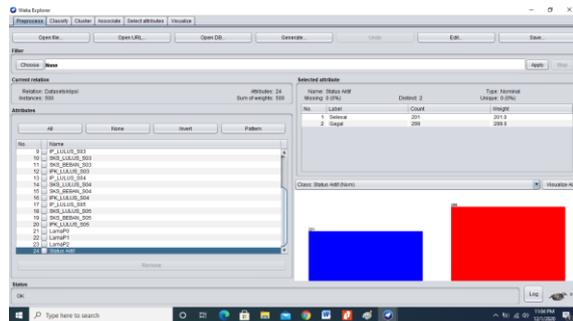


Gambar 4. Visualisasi SKS Lulus Semester 2

Gambar 4 menunjukkan hasil visualisasi dari atribut jumlah SKS Lulus Semester 2. Jumlah SKS lulus yang berupa data numerik ditransformasikan menjadi 5 kategori. Diketahui dari 500 data mahasiswa, ada 29 kategori Sangat Baik, 465 kategori Baik, 5 kategori cukup, 0 kategori kurang, dan 1 kategori sangat kurang. Gambar 5 menunjukkan hasil visualisasi dari atribut lama pengerjaan skripsi P1.



Gambar 5. Visualisasi dari Lama Pengerjaan Skripsi P1

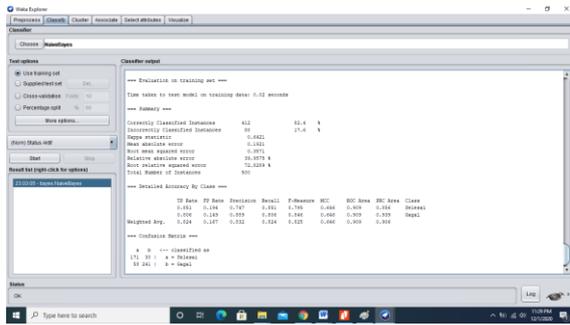


Gambar 6. Visualisasi dari Status Aktif

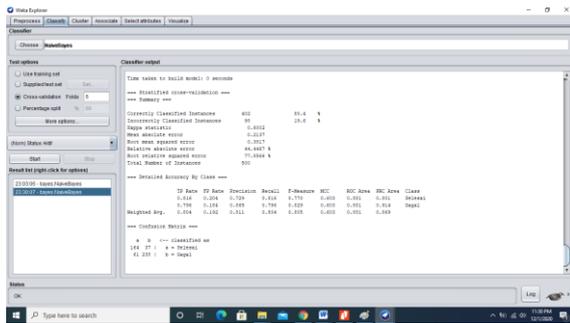
Lama pengerjaan skripsi P1 yang berupa data numerik ditransformasikan menjadi 2 kategori. Diketahui dari 500 data mahasiswa, ada 289 kategori Tepat Waktu dan 211 Tidak Tepat Waktu. Sedangkan Gambar 6 menunjukkan hasil visualisasi dari atribut status aktif. Status aktif menunjukkan status selesai skripsi atau gagal skripsi. Diketahui dari 500 data mahasiswa, ada 201 kategori Selesai dan 299 Gagal.

Pengujian implementasi algoritme naive bayes dilakukan dengan 3 set yaitu menggunakan *Use Data Training*, *5-cross validation*, dan *10-cross validation*. Hasil pengujian implementasi algoritme naive bayes menggunakan Weka dapat dilihat pada Gambar 7, 8, dan 9. Gambar 7 menunjukkan hasil implementasi naive bayes pada Tools Weka dengan menggunakan *use data training* yang menunjukkan hasil 412 prediksi benar dengan akurasi sebesar 82.4% dan 88 prediksi salah dengan presentasi sebesar 17.6%.

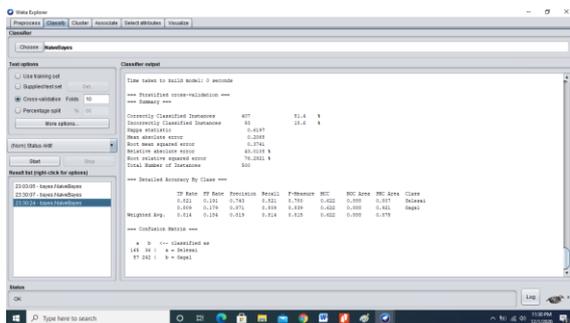
Gambar 8 menunjukkan hasil implementasi naive bayes pada Tools Weka dengan menggunakan *5-cross validation* yang menunjukkan hasil 402 prediksi benar dengan akurasi sebesar 80.4% dan 98 prediksi salah dengan presentasi sebesar 19.6%. Sedangkan Gambar 9 menunjukkan hasil implementasi naive bayes pada Tools Weka dengan menggunakan *10-cross validation* yang menunjukkan hasil 407 prediksi benar dengan akurasi sebesar 81.4% dan 93 prediksi salah dengan presentasi sebesar 18.6%.



Gambar 7. Hasil Naive Bayes (Use Data Training)



Gambar 8. Hasil Naive Bayes (5-cross validation)



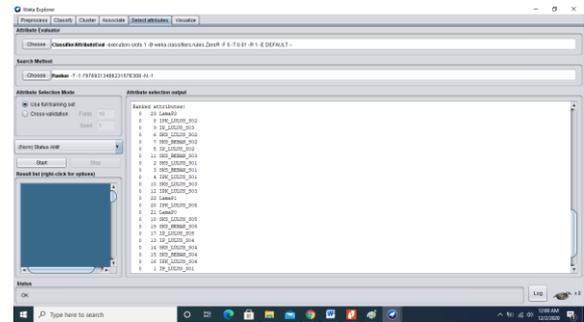
Gambar 9. Hasil Naive Bayes (10-cross validation)

Tabel 1. Perbandingan Evaluasi Akurasi

Model Satuan Evaluasi	Akurasi	Jumlah Kelas	Persentase
Use Data Training	Prediksi benar	412	82.4%
	Prediksi salah	88	17.6
5-cross validation	Prediksi benar	402	80.4%
	Prediksi salah	98	19.6%
10-cross validation	Prediksi benar	407	81.4%
	Prediksi salah	93	18.6%

Hasil perbandingan evaluasi akurasi dari data skripsi mahasiswa setelah dilakukan analisis implementasi algoritme *naive bayes*. Hasil perbandingan algoritme *naive bayes* pada tools Weka dengan menggunakan *Use Data Training*, *5-cross validation*, dan *10-cross validation* maka akurasi tertinggi yaitu algoritme *naive bayes* dengan menggunakan *use data training* yang menunjukkan hasil 412 prediksi benar dengan akurasi sebesar 82.4% dan 88 prediksi salah dengan presentasi sebesar 17.6%. Perbandingan hasil analisis dapat dilihat pada Tabel 1. Berdasarkan hasil seleksi atribut yang telah dilakukan menggunakan algoritme *classifier* maka atribut dengan *ranking* 10 teratas sangat mempengaruhi hasil klasifikasi. Atribut

tersebut yaitu atribut Lama P2, IPK Lulus Semester 2, IP Lulus Semester 3, SKS Lulus Semester 2, SKS Beban Semester 2, IP Lulus Semester 2, SKS Beban Semester 3, SKS Lulus Semester 1, SKS Beban Semester 1, dan IPK Lulus Semester 1. Lama P2 merupakan atribut dengan ranking teratas yang mempengaruhi hasil klasifikasi. Hasil seleksi atribut di Weka dapat dilihat pada Gambar 10.



Gambar 10. Hasil Seleksi Atribut di Weka

#### 4. KESIMPULAN DAN SARAN

Kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan dengan pengujian implementasi algoritme *naive bayes* dilakukan dengan 3 set yaitu menggunakan *Use Data Training*, *5-cross validation*, dan *10-cross validation*. Akurasi tertinggi yaitu algoritme *naive bayes* dengan menggunakan *use data training* yang menunjukkan hasil 412 prediksi benar dengan akurasi sebesar 82.4% dan 88 prediksi salah dengan presentasi sebesar 17.6%. Sedangkan hasil seleksi atribut yang telah dilakukan terdapat atribut dengan ranking 10 teratas yang sangat mempengaruhi hasil klasifikasi. Atribut tersebut yaitu atribut Lama P2, IPK Lulus Semester 2, IP Lulus Semester 3, SKS Lulus Semester 2, SKS Beban Semester 2, IP Lulus Semester 2, SKS Beban Semester 3, SKS Lulus Semester 1, SKS Beban Semester 1, dan IPK Lulus Semester 1. Lama P2 merupakan atribut dengan ranking teratas yang mempengaruhi hasil klasifikasi.

Saran penelitian ini dapat dilanjutkan pengembangan sistem cerdas untuk memprediksi kegagalan mahasiswa dalam mengerjakan skripsi. Sistem cerdas dibangun dapat menggunakan 10 atribut *Ranking Teratas*. Saran yang berikutnya juga dapat melakan analisis seleksi atribut non-akademik untuk memprediksi kegagalan skripsi. Karena penelitian ini hanya menggunakan atribut-atribut akademik.

#### DAFTAR PUSTAKA

ASNAWI, M. H., ISAL FIRMANSYAH, RAFLY NOVIAN, DAN RESA SEPTIANI PONTOH. 2021. Perbandingan Algoritme Naive Bayes, K-NN, dan SVM dalam pengklasifikasian Sentimen Media Sosial. Seminar Nasional Statistika X, vol. 10

- HERLAMBANG, A. D. DAN SATRIO H. W. 2019. Algoritme Naïve Bayes Untuk Klasifikasi Sumber Belajar Berbasis Teks Pada Mata Pelajaran Produktif Di Smk Rumpun Teknologi Informasi Dan Komunikasi. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* , vol. 6, no. 4, pp. 431-436.
- HERLAMBANG, A. D., SATRIO H. W., DAN ADITYA R. 2019. Intelligent Computing System to Predict Vocational High School Student Learning Achievement Using Naïve Bayes Algorithm. *Journal of Information Technology and Computer Science*, vol. 4, no. 1, Juni 2019. pp.15–25.
- NURAENI, R., ASO SUDIARJO, & RANDI RIZAL. 2021. Perbandingan Algoritme Naïve Bayes Classifier, dan Algoritme Decision Tree Untuk Analisa Sistem Klasifikasi Judul Skripsi. *Innovation in Research of Informatics (IINOVATICS)*, vol 3, No. 1, pp.26-31
- ROMERO, C., & VENTURA, S. 2010. Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601–618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- SHAKEEL, K., & ANWER BUTT, N. 2015. Educational Data Mining to Reduce Student Dropout Rate by Using Classification. 253rd OMICS International Conference on Big Data Analysis & Data Mining, (May). Retrieved from [https://www.researchgate.net/publication/281149091\\_Educational\\_Data\\_Mining\\_to\\_Reduce\\_Student\\_Dropout\\_Rate\\_by\\_Using\\_Classification?enrichId=rgreq95ac86e292913b9c0a37d546c75096e5XXX&enrichSource=Y292ZXJQYWdlOzI4MTE0OTA5MTtBUzoYnJyNzk2MzI1NzI0MTZAMTQ0MDI1](https://www.researchgate.net/publication/281149091_Educational_Data_Mining_to_Reduce_Student_Dropout_Rate_by_Using_Classification?enrichId=rgreq95ac86e292913b9c0a37d546c75096e5XXX&enrichSource=Y292ZXJQYWdlOzI4MTE0OTA5MTtBUzoYnJyNzk2MzI1NzI0MTZAMTQ0MDI1)
- YUAN, L. 2010. An Improved Naive Bayes Text Classification Algorithm in Chinese Information Processing. In: *Proceedings of the Third International Symposium on Computer Science and Computational Technology (ISCSCT '10)*, pp. 267-269. Jiaozuo, P. R. China
- WICAKSONO, A., H., AHMAD A. S., DAN SATRIO H. W., 2022. Klasifikasi Siswa Slow Learner untuk Mendukung Sekolah dalam Meningkatkan Pemahaman Siswa Menggunakan Algoritme Naïve Bayes. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*. Vol. 9, No. 3, Juni 2022, hlm. 589-596